DOCUMENT RESUME

ED 242 742 TM 840 122

AUTHOR Dorr-Bremme, Donald W.
TITLE Assessing Students: Teachers' Routine Practices and Reasoning.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Oct 83
NOTE 13p.
PUB TYPE Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT Evaluation Comment; v6 n4 p1-12 Oct 1983

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Criterion Referenced Tests; *Decision Making; Elementary School Teachers; Elementary Secondary Education; Grouping (Instructional Purposes); Norm Referenced Tests; Secondary School Teachers; *Standardized Tests; Student Evaluation; Student Placement; *Teacher Attitudes; *Teacher Made Tests; *Test Results; *Test Use

ABSTRACT

Some of the major findings of CSE's (Center for the Study of Evaluation) Test Use in Schools Project are synthesized and interpreted. The Project incorporated fieldwork and survey techniques to answer questions about the kinds of tests teachers administer in their classrooms, the kinds of information teachers need from tests to make decisions about their students, and how teachers use test information to make decisions. Data collected during the study are described and interpreted from the standpoint of teachers' routine assessment needs and practices. The classroom teacher is seen as a practical reasoner and decision maker who makes clinical use of assessment information to diagnose, prescribe, and monitor instruction. The tests teachers use most frequently are those that fit their practical circumstances: formal and informal measures they themselves construct or seek out for the information they provide; and curriculum embedded tests that come with commercial or district materials. Policy implications germane to the development of testing programs are presented and features of a testing system that could be directly useful to teachers are described. (LC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Center for the Study of Evaluation

October 1983 Vol.6, No.4

CSE EVALUATION COMMENT UCLA

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED242742

In This Issue . . .

Some of the major findings of CSE's Test Use in Schools Project are synthesized and interpreted. CSE's nation-wide study, funded by the National Institute of Education and conducted between 1979 through 1982, incorporated fieldwork and survey techniques to answer such questions as: what kinds of tests and other assessment devices do teachers *administer* in their classrooms? what kinds of information do teachers need from the tests and other devices they *use* to make decisions about their students? *how* do teachers use the information as they make these decisions?

Don Dorr-Bremme, a Senior Research Associate at CSE, provides answers to these and related questions. He describes the data collected during the study and interprets them from the standpoint of teachers' routine assessment needs and practices. In this interpretation, which draws on ethnomethodological concepts and sociological studies of professional groups, the classroom teacher is seen as a practical reasoner and decision maker who makes clinical use of assessment information to diagnose, prescribe, and monitor instruction.

From this data-based interpretation, Dorr-Bremme presents some policy implications germane to the development of testing programs. He describes some of the features of a testing system which, in addition to any broader applications that may be intended, can be directly useful for teachers as they go about the business of providing instruction and finding out how well their students have learned.

Assessing Students: Teachers' Routine Practices and Reasoning

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Donald W. Dorr-Bremme

TM 820122

Introduction

In the 1980's, testing issues confront educational policymakers at all organizational levels. The nation's investment in school achievement testing is enormous, and the amount and variety of testing continues to grow. Public accountability demands, mandates for minimum competency or proficiency testing, evaluation requirements for government-funded education programs, and judicial decisions defining increased responsibilities for public schools are only some of the factors that have fueled widespread debate about the nature and purposes of testing in the schools.

The quality of available tests has become a matter of controversy (CSE, 1979; The Huron Institute, 1978). Critics have indicted the validity of tests and attacked them as biased (Cabello, in press; Perrone, 1978). They have decried the arbitrariness of current testing practices (Baker, 1978), accused testing of narrowing the curriculum, and questioned the value of today's tests for the changing functions of American education (Tyler, 1977). Professional and advocacy groups representing educators, parents, and students have taken positions on testing. At least one major teacher's organization, for example, called for a moratorium on the use of standardized tests.

In response to these challenges, advocates of testing have asserted that tests can and do serve a variety of important purposes. They have maintained that achievement testing promotes high standards for learning, facilitates more accurate placement decisions, yields information for the improvement of curriculum and instruction, and helps the public hold schools accountable.

But as testing has proliferated and controversy has grown, little empirical information has been available on the assessment of student achievement as it is actually practiced in American classrooms. A small number of studies have indicated teachers' circumspect attitudes toward and limited use of one type of achievement measure—the norm-referenced, standardized test (e.g., Airasian, 1979; Boyd, et al., 1975; Goslin, Epstein, & Hilloch, 1965; Resnick, 1981; Salmon-Cox, 1981; Stetz & Beck, 1979). Another small body of research has suggested that students' social performances in school settings figure

Center for the Study of Evaluation
UCLA

Eva L. Baker
Director

EVALUATION COMMENT is published by the Center for the Study of Evaluation, University of California, Los Angeles. CSE is the only organized federally-supported agency devoting its total scope of work to the exploration and refinement of strategies in evaluation. CSE was established by the U.S. Office of Education in 1966 and is currently supported by the National Institute of Education. For 17 years, CSE has furnished the educational community with products and research which meet the increasing requirement for the evaluation of educational and social action programs.

Each issue of **Evaluation Comment** will discuss topics in educational evaluation by presenting articles on evaluation theory, procedures, methodologies, or practices. For each issue of **Comment**, CSE will request contributions on a specific topic in evaluation from recognized experts in the field; unsolicited manuscripts will not be accepted.

Each person on our mailing list receives a free copy of **Comment**. Where additional copies are needed readers are encouraged to reproduce **Comment** themselves. To be placed on our mailing list, please write to:

James Burry, Managing Editor
Evaluation Comment
145 Moore Hall
University of California, Los Angeles
Los Angeles, California 90024

Editorial Board

Eva L. Baker	Chairperson
Thomas Arciniega	Vice President for Academic Affairs California State University, Fresno
John B. Peper	Superintendent, Jefferson County School District Colorado
Michael Scriven	Director, Evaluation Institute University of San Francisco
Robert E. Stake	Director, CIRCE, and Professor of Educational Psychology at the University of Illinois
Daniel L. Stufflebeam	Professor of Educational Leadership and Director of the Evaluation Center, Western Michigan University

significantly in educators' judgments of students' scholastic competence (e.g., Cicourel & Kitsuse, 1963; Erickson & Shultz, 1982; Leiter, 1974; Rist, 1970). But studies such as these have offered only brief glimpses of teachers' reasoning and practices in evaluating student achievement. What methods and instruments do teachers routinely employ in making sense of how their students are doing academically? How do teachers think and reason about assessing their students' learning? Such questions as these have gone largely unaddressed.

In this context, the Center for the Study of Evaluation's (CSE) Test Use in Schools Project has begun to provide basic, new information on classroom achievement assessment across the United States. This article draws upon findings from the project's various phases, presenting an interpretive summary of some practices and kinds of reasoning which teachers routinely employ as they evaluate their students' achievement in the basic skills. First, however, the research itself will be briefly reviewed.

An Overview of the Test Use in Schools Project

Conducted from 1979 through 1982 (with some secondary data analysis still underway), CSE's test use research proceeded from broad definitions of *test* and *testing*. It encompassed a wide range of types of formal assessment measures, including commercially produced norm- and criterion-referenced measures; tests of minimal competency or functional literacy; and district-, school-, and teacher-constructed tests. Less formal measures for gauging student achievement, such as teachers' observations of and interactions with learners, were included as well. Within this broad domain, inquiry focused on achievement assessment practices and uses in Reading/English and Mathematics as carried out in public schools at the upper-elementary and high school levels.

During the project's first year of exploration and planning, comprehensive, semi-structured interviews were conducted in nine schools, three each in three school districts located in different states and geographic regions of the country. The districts and the elementary and secondary schools visited varied in size and demographic setting. Each of the interviews lasted about an hour and focused on assessment practices and uses of test results in the basic skills subjects mentioned above. Included among the interview respondents were 44 classroom teachers (22 elementary and 22 secondary), as well as principals, department chairpersons, counselors, and instructional specialists. Their remarks were tape recorded, transcribed, and coded using inductively developed categories.

Also during the first year, data collected in an earlier CSE study of testing and test use (Yeh, 1978) were reanalyzed. These data were gathered by self-administered questionnaires in 19 schools in five California school districts; some 256 teachers in grades K-6 responded.

A literature review accompanied the two exploratory research efforts in the project's first year.

In the project's second year, these activities informed the design of a nation-wide survey of teachers and principals. Survey recipients were drawn in a multi-step process. First, a nationally representative sample of some 114 districts was selected. This sample was stratified on the basis of district socio-economic status, enrollment size, minimum-competency testing policy, urban-suburban-rural locale, and geographic region of the country. From within these districts, size permitting, two elementary and two high schools were randomly chosen using a procedure that facilitated (where possible) inclusion of schools serving both higher- and lower-income populations. Finally, in each of these schools, principals received directions for randomly drawing four teachers for inclusion in the study. The principal and each of the four participating teachers received questionnaires that elicited detailed information on individual and school assessment practices as well as on related contextual and attitudinal data.

Returns were obtained from 220 principals, 486 upper-elementary teachers, and 365 high school English and math teachers. (The overall rate of return was 54% of the desired 2,000 respondents.) To correct for differential rates by stratification sampling cells and to approximate a na-

tionally representative distribution of respondents, weightings were applied in all analyses.

Finally, while the first two years of the project focused on testing practices and the uses of assessment results, the third year concentrated on testing costs. Case study research produced portraits of the direct and indirect costs of basic skills testing in two school districts (one large and urban, another small and suburban), including detailed descriptions of the testing costs in one elementary school within each district.

Major findings from the Test Use in Schools Project are available in a number of technical reports and articles. In addition to the Yeh (1978) study cited earlier, Lazar-Morrison and others (1980) have presented the main themes identified in the review of test use literature. Major findings of the exploratory, first year interviews have been discussed by Burry, et al. (1981). Burry, et al. (1982) and Herman and Dorr-Bremme (1983) have described and analyzed some of the early survey results. And the case studies of testing costs appear in Dorr-Bremme, et al. (1983).

This article elaborates on these early project reports by synthesizing and interpreting findings from the reanalysis of Yeh's (1978) data, from the project's first year interviews, and from the national survey. It draws on concepts from ethnomethodology (Garfinkel, 1967; Mehan & Wood, 1975) and from sociological studies of professional groups to describe and analyze how teachers routinely think about and carry out the assessment of student achievement.

The Findings: How Teachers Routinely Think and Act in Assessing Student Achievement

How do teachers routinely think and act in assessing student achievement? In answer to that question, the findings of the CSE Test Use Project suggest that teachers think and act both as practical reasoners and decision makers and as clinicians. That is, as they go about the business of determining how the students in their class(es) are doing academically:

- They orient their assessment activities to the practical tasks they have to accomplish in their everyday routines and do so in light of the practical contingencies and exigencies that they face on the job.
- And as they do, they make sense of students' academic performances clinically. They take into acount all the "data" at hand "in this particular situation." Then they interpret that data based on what "everyone" who is a member of the world of educational practice knows about what things mean and how things work in classrooms.

That teachers do think and act in these ways when they are carrying out student assessment is evident in the following Test Use Project findings.

(1) *In interviews, teachers report their uses of test results as serving most heavily the functions that are most central to teaching-as-practiced.*

In the on-site interviews, teachers were able to describe with minimal constraints how they used test results and information from other assessment techniques. *The purposes they most frequently cited were those that constitute their most essential, routine work*: deciding what to teach and how to teach it to students of different achievement levels; keeping track of how students are progressing and how they (the teachers) can appropriately adjust their teaching; and evaluating and grading students on their performance (see Table 1). Clearly, these are the day-to-day routines of teaching.

Less frequently, respondents mentioned using assess-

Table 1

Types of Tests and the Uses of Their Results (Interview Data)

(Cells show the number of times the 44 interviewed teachers freely cited each use for each type of test)

TEST TYPES / Uses	Standardized	Curriculum-Embedded	District Objectives-Based	Minimum Competency	Statewide Assessment	School, Department Grade-Level	Teacher-Constructed	Diagnostic (Skills)	Other (Informal)	TOTAL
Planning Instruction	13	10	3	4	2	3	24	2	21	82
Referral/Placement	11	0	0	1	0	2	3	0	6	23
Within Classroom Grouping & Individual Placement	4	18	5	3	1	4	6	6	14	61
Holding Students Accountable for Work, Discipline	0	3	0	0	0	0	8	0	2	13
Assigning Grades	1	17	1	1	0	5	32	1	8	66
Monitoring Students' Progress	0	14	4	0	0	2	18	1	12	51
Counseling & Guiding Students	3	0	2	0	0	0	10	1	6	22
Informing Parents	0	0	1	0	0	0	0	0	1	2
Reporting to District Officials, School Board, etc.	0	1	2	0	0	0	0	0	3	6
Comparing Groups of Students, Schools, etc.	1	0	1	0	0	0	0	0	1	3
Certifying Minimum Competency	0	0	0	1	0	0	0	0	0	1
Total Use	33	63	19	10	3	16	101	11	74	330
Explicit Statements of Non-use	10	0	0	2	7	1	0	0	1	21
TOTAL CITATIONS	43	63	19	12	10	17	101	11	75	351

ment results in deciding to refer students who need special instruction and to counsel, advise, and direct students. These are important teaching responsibilities, but ones that serve to support or facilitate more basic instructional work.

Use of test results in such tasks as comparing groups of students and reporting to people at higher levels of the school and district organizational hierarchy were rarely mentioned by teachers. These uses of test results are not in themselves unimportant. The reporting of scores to the school board, for instance, may be of considerable moment for superintendents or even principals. Comparing achievement across classrooms or schools is of central concern to district administrators and program coordinators. And these reports and comparisons may ultimately affect teachers' daily professional lives. It is not that these activities are inherently trivial, then, that makes them nonsalient for teachers; it is their remoteness from teachers' routine tasks that makes them so.

(2) *The means of assessment on which most teachers rely most heavily are those which facilitate the accomplishment of their routine activities under the exigencies they face.*

Reanalysis of data from the earlier CSE test use study (Yeh, 1978) found among 256 elementary school teachers surveyed that of all the tests they gave to their students, *teacher-made tests* figured more heavily than others in teachers' classroom decision-making. The reanalysis also discovered that for assessing student progress teachers relied heavily on *interactions with and observations of students*.

On-site interviews supported and elaborated these findings. The 44 teachers interviewed volunteered (collectively) 351 uses for nine types of assessment techniques. (Again, refer to Table 1.) They reported more uses (101) and more kinds of uses for their own, *self-constructed tests and major assignments*, e.g., essays, reports, etc., than for any other assessment type. Uses for *other, less formal, teacher-developed strategies*—peer evaluations, oral exercises, conferences with students, consultations with students' former teachers, etc.—were mentioned next most frequently (74 times), followed by *curriculum-embedded tests available commercially or constructed by the local school districts* (63 times). Furthermore, for schools in each of the three districts studied, the aforementioned types of assessment were those in which students spent the greatest proportion of their total assessment time.

National survey results dramatically confirmed the generality of Yeh's (1978) and the project's first-year, fieldwork findings for both elementary and secondary teachers. Teachers were asked to rate information from various sources (tests and others) as crucial, important, somewhat important, unimportant, or not available for conducting four routine decision-making activities—initial planning, initial student grouping, grouping changes, and giving grades. For initially grouping or placing students in a curriculum, for changing students from one group or curriculum to another, and for assigning grades, *nearly every survey respondent reported that "my own observations and students' classwork" was a crucial or important source of information.* (See Tables 2 and 3.) The great majority of respondents also indicated that the results of the *tests they themselves developed* also figured as

Table 2

Elementary Teacher Uses of Assessment Information for Different Decision-making Purposes

(Percentages of teachers surveyed reporting use of this information as crucial or important for the specified purpose)

	Planning Teaching at Beginning of School Year		Initial Grouping or Placement of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
Source/Kind of Information	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Previous teachers' comments, reports, grades	57	52	62	55	x	x	x	x
Students' standardized test scores	57	54	57	52	55	53	17	16
Students' scores on district continuum or minimum competency tests	51	47	50	45	45	39	20	18
My previous teaching experience	94	94	x	x	x	x	x	x
Results of tests included with curriculum being used	x	x	78	67	83	82	75	77
Results of other special placement tests	x	x	61	56	x	x	x	x
Results of special tests developed or chosen by my school	x	x	x	x	58	52	42	42
Results of tests I make up	x	x	80	86	78	85	92	95
My own observations and students' classroom work	x	x	96	97	99	99	98	98

Table 3

High School Teacher Uses of Assessment Information for Different Decision-making Purposes
(Percentages of teachers surveyed reporting use of this information as crucial or important for the specified purpose)

	Planning Teaching at Beginning of School Year		Initial Grouping or Placement of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
Source/Kind of Information	English	Math	English	Math	English	Math	English	Math
Previous teachers' comments, reports, grades	28	29	34	40	x	x	x	x
Students' standardized test scores	47	29	49	30	62	39	12	8
Students' scores on district continuum or minimum competency tests	48	30	47	36	53	36	9	5
My previous teaching experience	99	97	x	x	x	x	x	x
Results of tests included with curriculum being used	x	x	45	35	58	43	44	31
Results of other special placement tests	x	x	42	26	x	x	x	x
Results of special tests developed or chosen by my school	x	x	x	x	50	31	28	34
Results of tests I make up	x	x	87	77	92	91	99	99
My own observations and students' classroom work	x	x	99	93	99	97	99	95

crucial or important in these same decisions. And many elementary school teachers also responded that the "*results of tests included with the curriculum being used*" figured heavily in their planning for their teaching and in placing and changing the placement of students. Far lower percentages of teachers rated the other types of information listed as crucial and important in carrying out any of the three activities.

Looking over all these findings, it is evident that the types of assessment that most teachers rely upon most heavily have three characteristics in common:

- *Immediate accessibility: teachers can give them when they choose and see the results promptly.*
- *Proximity between their intended purposes and teachers' practical activities.*
- *Consonance, from teachers' perspectives, between the content they cover and the content taught.*

Each of these features responds to the exigencies of teachers' practical circumstances.

Teachers must accomplish their instructional work—initial planning, distributing students, teaching, continued planning, evaluating—within a temporal structure to which are attached normative expectations. Teaching units, marking periods, semesters, school years—these and other divisions of school time each have inherent points of closure. By those end-points, given amounts of learning are expected to be accomplished. Thus, time presses; teachers and their students must "progress;" decisions most often cannot wait (c.f. Jackson, 1968; Sarason, 1971; Smith & Geoffrey, 1968).

Not only is teaching time rapidly moving, it is also very full. Teachers interviewed during both the exploratory fieldwork and the third-year costs study were asked to detail the time they spent on various job-related activities in a normal school week. When their estimates were aggregated, elementary teachers' estimates averaged 357 hours a year spent outside the classroom, or about nine hours each week during the school year. High school teachers, on the average, seemed to be spending 600 hours a year, or about 15 hours a week on job-related tasks outside the classroom. And, of course, classroom time itself is constantly busy. Thus, teachers use means of assessment that are immediately accessible—that can be employed at the appropriate moment in the flow of on-going instruction—and for which results are quickly available.

Teachers also operate in an environment of accountability and concern. The decisions that they make matter, in varying degrees, to students' educational futures and life chances. Minimum competency laws, as well as court suits filed for "failure to educate," testify to the social pressures that bear upon teachers. That teachers recognize these pressures and strive to act with consonant concern and effort is evident (e.g., Lortie, 1975). Thus, teachers use assessment techniques that in their perspective accurately measure what has been taught, that measure the effects of the instruction that they believe they have given. And in response to both time and accountability demands, as well as to their own concern with assessing accurately, they employ measures that they believe match with the routine activities that they, as classroom teachers, must accomplish.

Teachers' determination of which particular tests or other assessment techniques meet these last two criteria is routinely a practical matter, not a "scientific" or technical one. That is, teachers tend to use and consult the results of whatever measures are present in the setting and purportedly relevant for the purposes at hand. If such a test is unavailable and a practical need is perceived for one, teachers feel competent to construct it. The appropriateness of these procedures is continually reaffirmed "reflexively" (Mehan & Wood, 1975, p. 8ff.) in the recurrent interactional activities of everyone involved in the world of schooling.

Throughout the interviews conducted in the first and third years of research, teacher comments on the technical properties of tests were, with only a handful of exceptions, notably absent. Routinely present were remarks which took for granted the technical adequacy of tests and, simultaneously, treated their practical features as matters of primary interest. Thus, both the reanalysis of Yeh's (1978) data and the fieldwork found that teachers frequently use whatever tests come with their curriculum for placement in that curriculum. Similarly, they most often employ self-constructed and curriculum-embedded unit tests for assessing performance on a unit and (ultimately) for grading students. The exploratory on-site visits also discovered heavy use by instructional specialists (remedial reading teachers, teachers of the learning disabled, etc.) of readily available normed diagnostic tests, e.g., the Sucher-Allred Reading Placement Inventory and the Bergantz Inventory of Basic Skills, for diagnosing individual learning problems and developing individualized programs. That such tests were labeled as appropriate for use in these tasks led to their use in accomplishing these tasks. And simultaneously and reflexively, their use in accomplishing these tasks reaffirmed the appropriateness of their labels.[1]

In summary, the assessment techniques teachers use most—teacher-made tests and assignments, curriculum-embedded tests, and especially the phenomenological data on students' performance that teachers gather everyday in the classroom—respond to the practical exigencies teachers face and the routine tasks they must accomplish. And in their use of these means of evaluating student achievement, as well as in their selection of particular measures, teachers reveal themselves as practical reasoners and decision makers in their everyday professional lives.

(3) *When test results are differentially important for teachers, their importance varies with their responsiveness to the practical exigencies that surround the task at hand.*

As Tables 2 and 3 display, teachers rarely find standardized test results important in deciding on students' report card grades. Substantially greater proportions of teachers, however, report that they do give these test results important consideration when it comes to planning their teaching at the beginning of the year. Standardized test scores also figure as crucial or important for many teachers as they go about the business of distributing and re-assigning students to instructional groups and curricula.

In the context of grading, standardized tests have characteristics that are exactly the opposite of those assessment results that most teachers rely on most heavily. The classroom teachers interviewed, for instance, complained that standardized test scores for their current class(es) arrived in their hands too late in the school year to be of any use. In many cases, teachers never got them for the present year's students; their results arrived the following fall. Many interviewees also noted that the scores provided little diagnostic information; others pointed out that the content of such tests overlapped only partially with what they were teaching. As they are usually scheduled and employed, then, standardized tests *lack* immediacy of accessibility. Their purposes are not perceived as proximal to teachers' everyday tasks. (As one respondent put it, "they're for comparison, not diagnosis of my kids' weaknesses and strengths".) And many teachers perceived a poor fit between what they teach and what standardized tests cover.

Nevertheless, in the context of another activity, more teachers find results of standardized tests useful. At the beginning of the year, teachers can drop into the office and check the standardized test scores of their new class(es) as they plan what to teach and how to pace their teaching through the opening weeks of the year or semester. And where standardized tests scores are reported on the class rosters that teachers receive at the beginning of the new year, some teachers interviewed said that they skimmed the scores, noted those that deviated sharply from most other scores on the list, then visited counselors to check on the placement of the students in question. *Thus, depending on the context—i.e., on the activity at hand and the range of information available—the scores of a given type of test may or may not meet teachers' practical needs. In those contexts where they do, teachers take them into account. In those contexts where they do not, teachers generally disregard them.*

The points made in the foregoing discussion add further detail to the portrait of the teacher as practical reasoner and decision maker.

Given the way the teacher's everyday world is organized, standardized tests are often impractical as sources of information. The scores they provide cannot be used in the work that constitutes day-to-day teaching—tracking students' progress through units, adjusting instruction to fit on-going achievement, assigning grades, etc. But when practical circumstances allow and occasional practical needs arise, teachers do treat standardized test results as important information. Thus, viewed from within "the world known in common and taken for granted" by teachers, teachers' demeanor toward and actions regarding standardized test scores make practical sense.

(4) *Teachers' explicit comments on tests and testing orient to the routine tasks and practical circumstances of teaching.*

The above evidence substantiating the concept of the teacher as practical reasoner and decision maker is based on *what teachers say that they do* in using tests. A slightly different form of evidence—*what teachers report that they believe and think*—ratifies the same concept. In the fieldwork interviews, teachers' remarks repeatedly called attention to their need for tests that are immediately accessible, that are consonant with the material taught, and that produce results that are of value in light of the routine tasks they confront everyday. The following quotations are illustrative of these points.

- The (standardized test) is almost useless in the spring, which is too bad, because I feel there is some valuable information there, progress and growth. But we get the score the last week of school.
- That computer-processed data (on district, objective-based tests) can really be used with those kids that

[1] This is not to suggest that the educators in question never abandoned one test in favor of another. But when they did so, it appeared to be on practical grounds, i.e., when it was perceived that the test recurrently "didn't work" for accomplishing the task that had to be done.

need help. It does a better job of identifying students and student needs . . . I can now say 'the kid needs to work on objectives 2, 3, 5, and 9.'

- I don't feel we need to test, test, test; but if the information is something I can use to prescribe instruction, then I don't really mind giving it.
- In math, you know, it's a good idea to keep them (tests) in my class. As long as testing stays in math class it seems like it fits in, 'cause tests are part of taking math.
- In my class, I like to use the criterion-referenced tests of basic skills. The tests are geared to certain basic skills the book's developing—vocabulary, spelling, and writing.
- I don't use (the district reading tests) unless there are results that completely throw me—like someone who usually does a good job completely bombed. Then I'll do something about that, try to find some extra work to go over it.

The orientation to assessment "for all practical purposes" that emerges in these fieldwork interview remarks also appeared in the reanalysis of Yeh's (1978) data. There, on a five point rating scale where 5 = "very important," teachers rated the following considerations for selecting tests as high: test material is similar to what I presented in class ($\bar{X}$ = 4.5); the test has clear format ($\bar{X}$ = 4.4); the test is simple to administer and/or score ($\bar{X}$ = 4.2). These practical matters in test selection are consonant with patterns of teachers' concerns and actions as reported throughout this section.

Finally, a slightly different dimension of teachers' practical orientation to assessment appears in survey responses to attitude questions. On the survey questionnaires, large majorities of both elementary grade teachers (73%) and high school English and mathematics teachers (80% and 93%, respectively) agreed that "testing motivates my students to study."

(5) *For given activities and decisions, teachers most often use the results of various types of assessment techniques collectively in a "clinical" way.*

The on-site interviews indicated that teachers most often consider the results of several types of assessment techniques in carrying out a particular task. Of the 351 instances in which teachers interviewed cited their uses for particular test scores and other assessment results, in 237 cases the score and results were used as *one of many* information sources (see Table 4). Reanalysis of Yeh's (1978) research discovered the same phenomenon. In both pieces of planning research, it also became evident that teachers often revise decisions made on the basis of test scores in light of their ongoing experience with children in the classroom. Other research reports similar patterns of action by teachers (e.g., Airasian, 1979; Cicourel & Kitsuse, 1963; Leiter, 1974; Salmon-Cox, 1980; Shumsky & Mehan, 1974).

Once again, the results of the national survey substantiate these earlier findings. This is indicated in the distribution of survey responses to those questions that asked teachers to report on the importance of different types of assessment information. (Refer to Tables 5 and 6.) *Extremely high proportions of both elementary and secondary teachers reported giving at least some importance to each type of information listed under three of the decision-making activities previously discussed*: initial planning, initial grouping and placing of students for instruction, and reassignment of students to different groupings and curricula. One need not examine the response patterns of individual teachers, then, to ascertain that the vast majority of them take a wide variety of kinds of assessment information into account in making each of these three types of instructional decisions. A glance at Table 7 shows more. Not only do survey respondents indicate that they consult several sources of information on students' achievement in making a particular instructional decision, they also report thinking that many kinds of assessment techniques give them *crucial and/or important* information for that decision.

Put another way, it does not seem as if teachers base their decisions primarily on one kind of assessment information, then look to others merely for confirmation or the sake of form. Rather, they appear to weigh various kinds of data on student achievement collectively and to make sense of what it means more-or-less holistically. If this is the case, it is a practice typical of clinical professions. The sociologist Homans (1950) long ago pointed out:

> Clinical science is what a doctor uses at his patient's bedside. There, the doctor cannot afford to leave out of account anything in the patient's condition that he can see or test . . . It may be the clue to the complex . . . In action we must always be clinical. An analytical science is for understanding but not for action.

More recently Friedson (1970) has outlined other features of what he calls the "clinical mentality." Noting with Homans that the aim of the clinical practitioner "is not knowledge but action," Friedson goes on to point out that the clinician

> cannot suspend action in the absence of incontrovertible evidence or be skeptical of himself, his experience, his work and its fruit.

Table 4

Overall Patterns of Assessment Results Use: Interview Data

	Functional Importance					
	Sole Source of Information Consulted	One of Several Major Sources	One of Many Sources	Verification Source	Not Used	Total
Instances Mentioned by 44 Teachers	18 (5.1%)	65 (18.5%)	237 (67.5%)	10 (2.8%)	21 (6.0%)	351 (100%)

Table 5

Percentages of Elementary Teacher Respondents Indicating Use of Information as "Somewhat Important," "Important," or "Crucial" for Each Task

Source/Kind of Information	Planning Teaching at Beginning of School Year	Initial Grouping of Students	Changing a Student from One Group or Curriculum to Another	Deciding on Students' Report Card Grades
Previous teachers' comments, reports, grades	93	75	x	x
Students' standardized test scores	92	91	89	43
Students' scores on district continuum or minimum competency tests	92	91	90	55
My previous teaching experience	100	x	x	x
Results of tests included with curriculum being used	x	98	97	93
Results of other special placement tests	x	96	x	x
Results of special tests developed or chosen by my school	x	x	96	81
Results of tests I make up	x	96	97	99
My own observations and students' classroom work	x	99	100	100

Thus, Friedson continues, "the clinician is prone in time to trust his own *personal first-hand* experience" (c.f., Becker, et al., 1961) and to be "particularistic," emphasizing the uniqueness of individual cases. The clinical rationality, then, "is particularized and technical: it is a method of sorting the enormous mass of concrete data confronting [the practitioner] in individual cases" (Friedson, 1970, p. 171).

This same mentality is evident in teachers' routine reliance upon and primary trust in their personal, interactive experience with children in the classroom, as well as in their tendency to treat many types of assessment data as equally relevant. The clinical mentality is also evident in many interviewees' explanations of *why* the results of one test or one type of test—or even of tests in general—cannot be trusted without reference to everyday evidence.

- I don't rely heavily on a lot of the test scores because I find that . . . some students are test takers and others are not . . . some students can handle the format, the time limit, (but in many cases) students are capable of more than the test scores show.
- I hate to say it, but I'd say about a third of these students don't give it their best shot. They feel there's nothing in it for them. There's no grade for it; there's no use for it—so they don't care.
- If I see there are certain kids having trouble I may look at their folders and find out (more) about them. But I try not to be swayed by somebody else's judgment . . . I may get more out of them by what I'm telling them and trying to motivate them to do better than they've ever done before.
- You can't count on a score on one test too heavily. The kid could be sick or tired or just not feeling up to doing it that day. Maybe his parents had a fight the night before. Maybe he doesn't test well.

Numbers of other respondents voiced equivalent opinions.

Similar reasoning appeared when teachers' opinions of the factors which can influence test scores were elicited in a closed-end format in Yeh's (1978) questionnaire study. On a five point rating scale (where 5 = "great influence" on test scores), among the factors for which teachers rated influence as 3.0 or higher were the following: students' test-taking skills ($\bar{X}$ = 4.4); test directions, content, format, physical characteristics, student motivation ($\bar{X}$ = 4.3); unusual circumstances—special activities, distractions ($\bar{X}$ = 4.2); and parent interest ($\bar{X}$ = 3.0). Teachers' belief that particularistic features of the test, the testing situation, and the students can and do mediate test results and their appropriate interpretation is reflected again here.

Part of what "everyone knows" in the world of educational practice, then, is that students vary as test takers and that a variety of situational factors can influence students' test performance. Better, then, to rely on a variety of sources of information—especially one's day-to-day, first-hand observations of and interactions with the individual across a variety of recurrent performance settings in the classroom—and to make sense of all the data at hand "in this situation" in light of one's practical knowledge and one's clinical experience,[2] or so teachers apppear to reason.

Summary

A variety of routine tasks constitutes the world of teaching-as-practiced. Teachers must accomplish these tasks in a context characterized by recurrent time limits, others' demands for high performance and accountability at those deadlines, and teachers' own concerns with providing effective and appropriate instruction. These features of the world of teaching-as-practiced impinge upon teachers' testing practices and test use. Teachers' reasoning and decision making about assessment and its uses are structured by and oriented to their practical circumstances.

The purposes for which teachers use assessment results most often are those inherent in the most central activities of teaching as it is practiced: determining what to teach and how to teach it in general and to various class members in particular, determining from day to day whether what they teach is being learned and adjusting instruction as necessary to be sure it is; and giving students grades so that they and their parents will know how they are doing. The role of the teacher includes other, optional tasks and responsibilities. But especially in view of the current ethos of meeting individual learners' needs, the central activities cited above are the essential, constitutive activities of classroom teaching. For those purposes less intimately connected with the central work of teaching,

Table 6

Percentages of High School Teacher Respondents Indicating Use of Information as "Somewhat Important," "Important," or "Crucial" for Each Task

Source/Kind of Information	Planning Teaching at Beginning of School Year	Initial Grouping of Students	Changing a Student from One Group or Curriculum to Another	Deciding on Students' Report Card Grades
Previous teachers' comments, reports, grades	71	75	x	x
Students' standardized test scores	77	76	86	24
Students' scores on district continuum or minimum competency tests	78	78	83	26
My previous teaching experience	100	x	x	x
Results of tests included with curriculum being used	x	83	87	68
Results of other special placement tests	x	80	x	x
Results of special tests developed or chosen by my school	x	x	84	61
Results of tests I make up	x	97	98	100
My own observations and students' classroom work	x	99	100	99

[2]Perhaps the data and the analysis presented here explain why an overwhelming percentage of teacher survey respondents, at both the elementary and secondary levels, agree that minimum competency tests should be required of all students for promotion at certain grade levels or for high school graduation (in agreement: elementary teachers, 58%; high school English and math teachers, 86% and 90% respectively), while simultaneously agreeing that teachers should not be held accountable for students' scores on minimum competency or standardized achievement tests. (Agreeing that teachers should not be held accountable: 75% of the elementary school teacher respondents and 61% of both high school English and high school math teachers.)

use of assessment results seems to occur less frequently. Action, in the "world known in common and taken for granted" by teachers, is oriented to and constrained by the organizational features of daily classroom teaching.

The tests teachers use most frequently are those that fit their practical circumstances: formal and informal measures they themselves construct or seek out for the information they provide; curriculum-embedded tests that come with commercial or district materials. These are immediately accessible, proximal in intended purpose to the tasks teachers must accomplish, and content-consonant

Table 7

Percentages of Teachers Who Report Considering Many Types of Assessment Information Critical/Important for Given Activities

	Planning Teaching at Beginning of School Year	Initial Grouping or Placement of Students	Changing Grouping or Placement	Deciding on Report Card Grades
Number of Sources of Information Given in Question on Survey	4	7	6	6
Number of Sources Defined as "Many" for Purposes of This Analysis	3	4	4	4
Proportion of *Elementary Teachers* Who Indicated That at Least This Many Sources Functioned as Critical and/or Important for the Given Activity	50%	71%	62%	40%
Proportion of *High School* Teachers Who Indicated That at Least This Many Sources Functioned as Critical and/or Important for the Given Activity	33%	47%	49%	20%

with the material taught. The further removed from these qualities that tests and testing features are, the less willing teachers are to give them or to consider their results as important.

The way in which teachers use tests follows from their practical understandings of the "scenic features" of their world. They recognize—tacitly in their actions and often explicitly in their words—that performance varies with context and that many "readings" of student achievement are better than few. Thus, they most often use results from many assessment types collectively to accomplish given purposes. Their immediate, recurring experience with children often overrides scores from paper-and-pencil instruments.

Teachers' comments about tests and testing confirm their orientation to the practical business of getting everyday tasks done in time and done well. They speak of the need to diagnose, prescribe, and assess efficiently and accurately. They talk of the need for test directions and formats that are clear. And they comment practically about the need to consider "extenuating circumstances," to pass on information "which is meaningful to everybody," and the like.

Some Implications for Local Policy and Practice

If testing programs are to be *useful for teachers and used in classrooms*, they must take into account teachers' routine thinking and practices in assessing students' achievement. American educational organizations (schools, school districts, etc.) have been called "loosely coupled systems" (c.f., Deal, 1979; Meyer & Rowan, 1978; Montjoy & O'Toole, 1979). Schooling in the United States has been described as "pre-industrial—a cottage industry" (Dawson, 1977). And teachers in classrooms have been likened to "street-level bureaucrats" (e.g., Weatherly & Lipsky, 1977). These similes call attention to the relative autonomy of the classroom teacher in a multi-leveled decision-making hierarchy—a hierarchy in which participants at each level have interests and concerns that only partially overlap, only sometimes coincide. In such a system, innovation—such as the introduction of a new testing program—tends to be more enduring not when it is imposed from the top down, not when it is generated from the bottom up, but when it is planned and implemented conjointly by participants at all levels (Berman & McLaughlin, 1978).

Conjointly planned programs should incorporate features which are important to teachers. Our results indicate that teachers favor tests that are:

(1) *proximal to the everyday instructional tasks teachers need to accomplish*: planning their teaching, diagnosing students' learning needs, monitoring their progress through the curriculum-as-taught, placing students in appropriate groupings and instructional programs, adjusting their teaching in light of students' progress, and informing parents and others about how students are doing;
(2) *consonant, from teachers' perspectives, with the curriculum that teachers are actually teaching*;
(3) *immediately accessible* to teachers, so that teachers can give them to students when the time seems appropriate and have results available promptly;
(4) *designed to include a variety of performance "contexts,"* i.e., different types of response formats and tasks.

Many districts' (and schools') testing programs fail to meet these criteria in one or more ways. When they do, they become simply an extra burden for teachers. Instructional time is taken up in testing, but there are few concomitant benefits for teachers or students. In other cases, districts (and sometimes schools) hope to meet the above criteria by developing sets of tests oriented to local curricular objectives. But the Test Use Project's interviews and fieldwork indicate that in many cases these objectives-based tests only *seem* to meet the criteria listed above. Thus, the experience of one district studied by the project may provide a useful example of how those criteria can be met.

A case in point. The mid-western district in question (enrollment about 5,000) did not have vast resources. Nevertheless, it involved teachers during the school year and especially during the summer in building curricula and tests to accompany them. Teachers participated in substantial numbers. (And at the elementary level, they were the leaders of cross-grade-level teaching teams—leaders chosen by their colleagues.)

The emphasis in these recurrent projects was on curricular objectives and instructional materials. An effort was made to select objectives and design materials that teachers found appealing and used. Repeated revisions of instructional materials and goals in light of teachers' criticisms were part of the process. Tests were designed to fit each curriculum—tests that met teachers' routine teaching needs. Thus, the curricular packages included placement tests, chapter and unit tests, and semester and end-of-year review tests or "finals." These tests were also revised in response to teachers' criticisms during the development process, which included as a final step using the curricula and tests in schools throughout the district on a pilot basis for a year.

The tests themselves were designed to be computer scored and analyzed, using computers that the district had originally purchased for computer-assisted instruction in the high school. Teachers gave the tests at times they felt were appropriate, turned them in for scoring, and received the analyzed results within a day or two. The results themselves came in the form of a set of sheets, one for each student. The sheet listed (1) each objective the test covered, (2) the number of items that assessed performance on each objective, and (3) the number of items that the student passed and missed on each objective. At the top of the sheet was a paragraph listing the main types of errors that the student had made and stating just what problems the student seemed to be having. This was based on an analysis of the questions missed and the incorrect items chosen.

Teachers reported that they and their colleagues routinely used these tests. And interview response patterns indicated that they spent less time designing, administering, and scoring their own tests than teachers in other districts visited. Interviewees stated explicitly that they used these tests (1) because they fit so well with what they were actually teaching, (2) because they could be used flexibly, e.g., at any time, with one child or an entire class, (3) because scores came back promptly, and (4) because the analyses summarized information in a way that gave them precise diagnoses they could act on in placing students, in deciding who needed additional help on what skills, etc. In fact, the only complaint teachers had was that all the tests were multiple choice tests. As one teacher put it, "that's a problem, 'cause sometimes you wonder whether they can apply the skills or ideas in another way."

In short, this district made considerable efforts to assure that its testing program was useful to and used by teachers. In so doing, its program for testing fulfilled three of the four criteria identified earlier: the tests were proximal, instructionally consonant, and immediate; but they had only one response mode. The program met district needs, too. Semester and end-of-year finals functioned to indicate strengths and weaknesses of the students in particular schools and in schools throughout the district from year to year. Thus, they served various evaluation and management functions.

Testing programs which take into account teachers' routine thinking and practices in assessing students' achievement can probably take many shapes. The case described here is only one example. But it should be clear that programs of testing that ignore how teachers think and act toward student assessment can result in inefficiency and teacher resentment.

References

Airasian, P. W. *The effects of standardized testing and test information on teachers' perceptions and practices.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

Baker, E. L. *Is something better than nothing? Metaphysical test design.* Paper presented at the 1978 CSE Measurement and Methodology Conference, Los Angeles, 1978.

Becker, H. S., et al. *Boys in white: Student culture in medical school.* Chicago: University of Chicago Press, 1961.

Berman, P., & McLaughlin, M. W. *Federal programs supporting educational change. Vol. III: Implementing and sustaining innovations.* R-1589/8-HEW. Santa Monica, CA: RAND Corporation, 1978.

Boyd, J., Jacobsen, K., McKenna, B. H., Stake, R. E., & Yashinsky, J. *A study of testing practices in the Royal Oak (Michigan) Public Schools.* Royal Oak, Michigan: Royal Oak Public Schools, 1975.

Burry, J., Dorr-Bremme, D., Herman, J. L., Lehman, J., & Yeh, J. *Teaching and testing: Allies or adversaries.* CSE Report No. 165. Los Angeles: UCLA Center for the Study of Evaluation, 1981.

Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. *Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs?* CSE Report No. 194. Los Angeles: UCLA Center for the Study of Evaluation, 1982.

Cabello, B. A description of analysis for the identification of potential sources of bias in dual language achievement tests. To appear in the *NABE Journal* (National Association for Bilingual Education), 1983.

Center for the Study of Evaluation. *CSE criterion-referenced test handbook.* Los Angeles: UCLA Center for the Study of Evaluation, 1979.

Cicourel, A. V., & Kitsuse, J. I. *The educational decision-makers.* Indianapolis: Bobbs-Merrill, 1963.

Dawson, J. E. Why do demonstration projects? *Anthropology and Education Quarterly,* 1977, 8(2), 95–105.

Deal, T. E. *Linkage and information use in educational organizations.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

Dorr-Bremme, D., Burry, J., Catterall, J., Cabello, B., & Daniels, L. *The costs of testing in American public schools.* CSE Report No. 198. Los Angeles: UCLA Center for the Study of Evaluation, 1983.

Erickson, F., & Shultz, J. *The counselor as gatekeeper: Social interaction in interviews.* New York: Academic Press, 1982.

Friedson, E. *Profession of medicine: A study of the sociology of applied knowledge.* New York: Dodd, Mead & Company, 1970.

Garfinkel, H. *Studies in ethonomethodology.* Englewood Cliffs, NJ: Prentice-Hall, 1967.

Goslin, D. A., Epstein, R., & Hilloch, B. A. *The use of standardized tests in elementary schools.* Second Technical Report. New York: The Russell Sage Foundation, 1965.

Herman, J. L., & Dorr-Bremme, D. Uses of testing in the schools: A national profile. To appear in W. Hathaway (Ed.), *New directions for testing and measurement: Testing in the schools.* San Francisco: Jossey Bass, Inc., 1983.

Homans, G. *The human group.* New York: Harcourt, Brace & World, 1950.

Huron Institute. *Summary of the spring conference of the National Consortium on Testing.* Cambridge, MA: Huron Institute, 1978.

Jackson, P. *Life in classrooms*. New York: Holt, Rinehard and Winston, 1968.

Lazar-Morrison, C., Polin, L., Moy, R., & Burry, J. *A review of the literature on test use*. CSE Report No. 144. Los Angeles: UCLA Center for the Study of Evaluation, 1980.

Leiter, K. C. W. Ad hocing in the schools: A study of placement practices in two kindergartens. In A. V. Cicourel, et al., *Language use and school performance*. New York: Academic Press, 1974.

Lortie, D. C. *Schoolteacher: A sociological study*. Chicago: The University of Chicago Press, 1975.

Mehan, H., & Wood, H. *The reality of ethnomethodology*. New York: Wiley Interscience, 1975.

Meyer, J. W., & Rowan, B. The structure of educational organizations. In M. W. Meyer and associates (Eds.), *Environment and organizations*. San Francisco: Jossey-Bass, 1978.

Montjoy, R. S., & O'Toole, Jr., L. F. Toward a theory of policy implementation. *Public Administration Review*, 1979, *39*(5).

Perrone, V. *Remarks to the National Conference on Achievement Testing and Basic Skills*. Paper presented at the National Conference on Achievement Testing and Basic Skills, Washington, D.C., 1978.

Resnick, L. B. Introduction: Research to inform a debate. *Phi Delta Kappan*, 1981, *62*(9), 623-625.

Rist, R. C. Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 1970, *40*(3), 411-450.

Salmon-Cox, L. Teachers and tests: What's really happening? *Phi Delta Kappan*, 1981, *69*(9), 631-634.

Sarason, S. *The culture of the school and the problem of change*. Boston; Allyn and Bacon, 1971.

Shumsky, M., & Mehan, H. *The structure and practices of description in two evaluative contexts*. Paper presented at the Eighth World Congress of Sociology, Toronto, Canada, 1974.

Smith, L., & Geoffrey, W. *The complexities of an urban classroom: An analysis toward a general theory of teaching*. New York: Holt, Rinehard & Winston, 1968.

Stetz, F., & Beck, M. *Comments from the classroom: Teachers' and students' opinions of achievement tests*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 1979.

Tyler, R. What's wrong with standardized testing? *Today's Education*, 1977.

Weatherly, R., & Lipsky, M. Street-level bureaucrats and institutional innovation. *Harvard Educational Review*, 1977, *47*(2).

Yeh, J. P. *Test use in schools*. Los Angeles: UCLA Center for the Study of Evaluation, 1978.

... Other Highlights from the Test Use Project ...

- Elementary teachers report that testing in reading takes up about 13 hours of student time each year, and that math testing takes up about 17 hours of student time annually.
- Secondary teachers report that testing in English and in math each take up about 25 hours of student time each year.
- Elementary school teachers report that about half their students' testing time is spent in testing required by local, state, or federal agencies; secondary teachers report that about one-quarter of their students' testing time is mandated.
- Teachers report that testing and test-related matters consume about 5-7 hours each week— about 12% to 15% of their total work effort during and after school hours.
- Case study findings indicate that testing, across all subjects, takes up about 10% of students' annual classroom time.
- A majority of teachers favor minimum competency testing for promotion and high school graduation but, particularly where these are required, are concerned about their fairness for some students.
- A majority of teachers feel that minimum competency tests affect the amount of time they can spend teaching subjects not covered by the tests.
- Most teachers report that they do not receive staff development training or other assistance in selecting or developing good tests or in using tests to improve instruction.
- Survey findings indicate that administrative support for testing, exemplified in principal interest, available resources, and/or opportunities for staff development training, influence teachers' attitudes toward and use of testing.

CSE and the UCLA research community are sad to report the death of Bruce Choppin, formerly director of the CSE methodology program. We extend our sympathy to family and friends.

We are also saddened by the death of Ronald Edmonds, of Michigan State University, who served as a visiting scholar to CSE this year.